5. STUDY DESIGN - PIVOTAL STUDIES

- Four, multicenter, randomized, double-blind, placebo-controlled, parallel-group studies (039, 042, 091, and 156) evaluated the use of cilomilast 15mg BID for 24-weeks in patients with COPD.
- Patients randomized into the studies had moderate to severe COPD with poor reversibility to albuterol.

5.1. Dose Rationale

Although a clear dose ordering of response was not established in the dose-response study (032), the magnitude of effect of cilomilast 15mg BID dose was greater than the 5mg or 10mg BID doses. Cilomilast 15mg BID was the only dose superior to placebo with regard to significant improvements in pulmonary function and positive trends in symptoms and quality of life in patients with COPD. Cilomilast 15mg BID as the dose of choice for further clinical trials was supported by two Clinical Pharmacology studies (003 and 140) where cilomilast administered as a single dose of 20mg or greater was poorly tolerated. Therefore, a 15mg BID dose was selected for evaluation in the Phase III clinical program.

5.2. Study Design

Four pivotal, multicenter, randomized, double-blind, placebo-controlled, parallel-group studies (039, 042, 091, and 156) evaluated the use of cilomilast for 24-weeks in patients with COPD. The primary objectives of the studies were to assess the clinical efficacy of oral cilomilast 15mg BID by assessment of trough FEV_1 and quality of life as measured by the St. George's Respiratory Questionnaire (SGRQ) over 24 weeks in patients with COPD. Secondary objectives included the assessment of additional efficacy parameters, including COPD exacerbations and to further define the clinical safety and tolerability of cilomilast.

In the pivotal studies, cilomilast 15mg BID was compared to placebo in a 1:1 (Study 156) or 2:1 (Studies 039, 042, and 091) ratio in populations ranging from 647 to 825 patients per study. Each study had a 4-week, single-blind, placebo run-in period followed by 24 weeks of double-blind treatment. One tablet of cilomilast 15mg or matched placebo was taken BID, immediately after breakfast and after the evening meal. The treatment period was followed by a 1-week safety follow-up period for patients who did not enter an open-label extension or who withdrew prior to the end of the study.

The following concomitant COPD medications were permitted during the studies: stable doses of anticholinergic medication via a metered dose inhaler (MDI), albuterol (on an "as needed" basis) via MDI and mucolytics. Additional COPD medications were allowed for <14 days to treat COPD exacerbations. Xanthines were not allowed at any point during the clinical trials.

Key inclusion and exclusion criteria were similar across the four pivotal studies and included the following:

5.2.1. Key Inclusion Criteria

- Male or female adults between 40 and 80 years of age
- Clinical diagnosis of COPD, as defined by ATS guidelines (for NA Studies 039 and 156) or European Respiratory Society (ERS) guidelines (for EU Studies 042 and 091)
- Cigarette smoking history of ≥10 pack years
- Pre-bronchodilator FEV_1 to FVC ratio ≤0.7 at Screening
- Fixed airway obstruction, defined as ≤15% or ≤200mL increase in FEV_1 (or both) post-bronchodilator at Screening
- Post-bronchodilator FEV_1 between ≥30% and ≤70% of predicted normal for height, age and sex at Screening [patients were assessed 15 to 30 (±5) minutes after receiving albuterol/salbutamol]

5.2.2. Key Exclusion Criteria

- Patients with asthma as the main component of their obstructive airways disease
- Patients with poorly controlled COPD (defined as the occurrence of any of the following in the 2 weeks prior to Screening: acute worsening of COPD that was managed by the patient at home by self-treatment with corticosteroids or antibiotics, that required treatment prescribed by a physician, or for which the patient was hospitalized)
- Patients with active tuberculosis, lung cancer, or clinically overt bronchiectasis
- Patients with clinically significant cardiovascular, neurological, renal, endocrine, or hematological abnormalities that were uncontrolled on permitted therapy
- Patients with clinically significant gastrointestinal or hepatic abnormalities
- Patients with a positive fecal occult blood test result between Screening and Baseline visits
- Patients with clinically significant orthostatic changes in blood pressure or heart rate at Screening or Baseline visits
- Patients who required treatment with inhaled cromolyn sodium or nedocromil, inhaled long-acting β2-agonists, oral β2-agonists, nebulized β2-agonists, nebulized anticholinergics, xanthines, leukotriene modifiers or oral/inhaled corticosteroids beyond Screening
- Patients receiving treatment with long-term oxygen therapy, patients who required supplemental oxygen more often than on an occasional basis, or patients who required nocturnal positive pressure for sleep apnea
- Patients who had participated in a Pulmonary Rehabilitation Program within 4 weeks prior to Screening or who planned to enter a Pulmonary Rehabilitation Program during the study

Across all four trials, the study populations had poor reversibility of airway obstruction, with a ≤15% or ≤200mL increase in FEV_1 (or both) post-bronchodilator at Screening. However, there were differences in the doses of bronchodilator (albuterol/salbutamol) used to measure reversibility in the four pivotal studies. EU Studies 042 and 091 utilized albuterol administered at a dose of 360mcg ex-actuator via MDI (i.e., 400mcg ex-valve) while NA Studies 039 and 156 utilized albuterol administered at a dose of 180mcg ex-actuator via MDI (i.e., 200mcg ex-valve).

Randomization Criteria

In order to be randomized at the end of the single-blind placebo run-in phase, patients were required to meet the following criteria:

- Stability in Pulmonary Function: Clinic trough FEV_1 (absolute value) in the absence of albuterol/salbutamol did not increase or decrease by more than 20% between Screening and Baseline.
- Diary Symptom Score: A total symptom score (cough + sputum production + breathlessness) of 3 or more was achieved on at least 5 of the 10 days immediately prior to the Baseline visit. (This criterion was not used in Study 156).

5.2.3. Efficacy Measures

Across the four pivotal studies, the co-primary endpoints were FEV_1 and total score of the SGRQ. Primary and secondary efficacy parameters for these studies are shown in Table 2.

Table 2 Efficacy assessments in the pivotal efficacy studies

Efficacy Parameters	Study 039	Study 156	Study 042	Study 091
Primary Efficacy Parameters				
FEV_1	X	X	X	X
SGRQ	X	X	X	X
Secondary Efficacy Parameters				
COPD exacerbations	X	X	X	X
FVC	X	X	X	X
Exercise tolerance (6-minute walk)	X	X[a]	X	X
Post-exercise breathlessness (modified Borg Scale)	X	X	X	X
Summary Symptom Score (comprising the sum of scores for cough, sputum and breathlessness recorded on domiciliary diary card)	X		X	X

a. Tertiary efficacy parameter in Study 156.

5.2.3.1. Study Procedures

Pulmonary Function Tests

Pulmonary function was assessed using standardized spirometry equipment supplied by GSK that met or exceeded the minimal performance recommendations of the ATS

[American Thoracic Society, 1995]. Pulmonary function tests were performed at trough study medication plasma levels. FEV_1 was measured in triplicate at each clinic visit at least 2 hours (4 hours for Study 156) after the patient's last use of β2-agonist or anticholinergic medication. For each set of FEV_1 measurements, one of three had to be within 10% of the highest measurement. The highest FEV_1 and forced vital capacity (FVC) measurements were recorded. Predicted FEV_1 was calculated as described by Crapo et al [Crapo, 1981].

SGRQ

The SGRQ is a self-administered questionnaire designed to measure and quantify the impact of chronic respiratory disease on health-related quality of life and well-being. The SGRQ measures the impact of chronic respiratory disease on three domains: symptoms, activity, and impact (on daily life). The first part of the questionnaire ("Symptoms") contains questions that evaluate distress due to respiratory symptoms, including frequency of cough, sputum production, wheeze, breathlessness, and the duration and frequency of attacks of breathlessness or wheeze. The second part has two components, "Activity" and "Impacts". The "Activity" section addresses the effects of disturbance due to mobility and physical activity that cause breathlessness or are limited because of breathlessness. The "Impacts" section covers a range of factors quantifying the psychosocial impact of the disease including influence on employment, being in control of health, panic, stigmatization, the need for medication, side effects of prescribed therapies, expectations for health, and disturbances of daily life.

A total score was calculated for the questionnaire and a 4-point change in score was considered to be a clinically meaningful difference [Jones, 1991]. A copy of the questionnaire and instructions for administering the questionnaire are included in the Appendices.

COPD Exacerbations

COPD exacerbations were categorized as Level 1, 2, or 3, based on the treatment received by the patient. The levels of exacerbation were defined as follows:

- Level 1 Exacerbation: Acute worsening of COPD that is self-managed by the patient at home by increasing usual COPD medications (i.e., the patient does not see a doctor for this episode).
- Level 2 Exacerbation: Acute worsening of COPD that requires additional treatment (e.g., a short course of oral steroids, antibiotics, etc.) prescribed by a family physician or primary care doctor or as a result of a hospital outpatient visit (including a visit to the Emergency Room).
- Level 3 Exacerbation: Acute worsening of COPD that requires the patient to be admitted to the hospital for treatment.

Summary Symptom Score

The summary symptom score is a summary of cough, sputum production, and breathlessness scores as recorded on a daily basis in patient diary cards. Patients were

instructed to complete the diary in the evening, and to base their assessment on symptoms experienced over the course of the day (24 hours). Patients assessed their symptoms using 0-4 point scale for breathlessness and a 0-3 point scale for cough and sputum production. The three scores were summed for a total summary symptom score.

Exercise Tolerance Test

Exercise tolerance was assessed using a 6-minute walk test to assess the functional capacity of the patient [Guyatt, 1985; Jones, 1988]. The patient was instructed and encouraged to walk from one end of a course to the other, and the distance walked was recorded.

Post-Exercise Breathlessness

Breathlessness following the exercise tolerance test was assessed using the modified Borg scale [Borg, 1982].

5.2.4. Safety Measures

Available safety data from the phase III pivotal studies included the following and more detailed information are presented in the corresponding sections:

Adverse events (Section 6.5.2)

Assessments of gastrointestinal safety (Section 6.5.2.4)

Clinical laboratory values (Section 6.5.3)

Sitting vital signs as well as orthostatic changes in vital signs (Section 6.5.4)

Electrocardiograms (Section 6.5.5.1)

24-hour Holter monitoring (Section 6.5.5.2)

5.2.5. Statistical Methods and Analyses

5.2.5.1. Sample Size

There were co-primary endpoints for the four pivotal studies. Sample size determinations for the pivotal studies were based on a standard deviation of 12 units in the change in the total score of the SGRQ to detect a difference of 4 units with 90% power at a significance level of 0.025 in Studies 039, 042, and 091, and a significance level of 0.05 in Study 156. Similarly with the same significance levels as for SGRQ, a standard deviation of 270mL in the change in FEV_1 was used to detect a difference of 120mL with greater than 90% power in studies 039, 042, and 091, and standard deviation of 210mL was used to detect a difference of 50mL with 90% power in Study 156, with correlation between visits of 0.68 for all four studies.

5.2.5.2. Populations for Analysis

Efficacy evaluations for the pivotal studies were made for the intent to treat (ITT) population, which includes all patients who received double-blind treatment and had at least one on-therapy FEV_1 assessment. An on-therapy efficacy evaluation was defined as an evaluation performed after the initial dose of randomized study medication at the scheduled visits and within 24 hours of the last dose. The ITT population was the primary population for analysis.

5.2.5.3. Statistical Analyses

Patient Disposition and Demography

The number and percent of patients who completed the study were tabulated. Demographic and baseline characteristics were provided for all randomized patients.

Primary Endpoints

Evidence for the efficacy of cilomilast in COPD was primarily provided by the statistical analyses performed individually for the four pivotal studies (039, 042, 091, and 156). Primary efficacy results were presented based on the average change from Baseline over the 24 week double-blind treatment period. Average change over the 24 weeks was estimated using a repeated measures analysis model with an unstructured covariance. For NA Studies 039 and 156, the repeated measures model included treatment, center, and time as fixed effects. For EU Studies 042 and 091, the model included treatment, country, and time as fixed effects.

Efficacy results were also presented based on the change from Baseline to Endpoint. Endpoint was defined as the last non-missing on-therapy assessment collected for a patient regardless of the type of visit at which the assessment was collected. Baseline assessments were not considered for Endpoint. Change from Baseline to Endpoint was analyzed using an analysis of variance (ANOVA) with treatment and center as fixed effects for the NA studies and treatment and country for the EU studies.

To account for co-primary endpoints in Studies 039, 042, and 091, the Hochberg method [Hochberg, 1988] was used to adjust the significance level in the test for treatment effect. This method proceeded by ordering the p-values from the analysis of each primary endpoint. If the larger p-value was less than 0.05, then both primary endpoints were declared significant. If the larger p-value was greater than or equal to 0.05, then the smaller p-value was compared against a significance level of 0.025. For Study 156, the p-values for both primary endpoints were compared against a significance level of 0.05 and significance was claimed only when both p-values were less than 0.05, therefore no adjustment of the significance level was needed in the test for treatment effect.

Secondary Endpoints

Results for post-exercise dyspnea, trough FVC, distance walked, and summary symptom score were based on the same ANOVA model as above at the Endpoint.

COPD exacerbation incidence over 24 weeks was estimated using the Kaplan-Meier product limit method. Relative risk of experiencing at least one COPD exacerbation was estimated using a Cox Proportional Hazards Model.

Safety

Safety data, including adverse event (AE), vital signs and laboratory tests, were displayed using descriptive statistics or graphs. No formal statistical comparisons were made for safety data.